



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### How do speakers coordinate?

**Citation for published version:**

Gambi, C, Cop, U & Pickering, MJ 2015, 'How do speakers coordinate? Evidence for prediction in a joint word-replacement task', *Cortex*, vol. 68, pp. 111-128. <https://doi.org/10.1016/j.cortex.2014.09.009>

**Digital Object Identifier (DOI):**

[10.1016/j.cortex.2014.09.009](https://doi.org/10.1016/j.cortex.2014.09.009)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Cortex

**Publisher Rights Statement:**

© Gambi, C., Cop, U., & Pickering, M. J. (2015). How do speakers coordinate?: Evidence for prediction in a joint word-replacement task. *Cortex*; a journal devoted to the study of the nervous system and behavior. [10.1016/j.cortex.2014.09.009](https://doi.org/10.1016/j.cortex.2014.09.009)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Word count (excluding references, footnotes, and appendices): 11038

How do speakers coordinate?

Evidence for prediction in a joint word-replacement task

Chiara Gambi<sup>a1</sup> Uschi Cop<sup>b</sup> Martin J. Pickering<sup>c</sup>

<sup>a</sup> Department of Psychology, University of Edinburgh, 7 George Square, EH8 9JZ,  
Edinburgh, U.K.; email: gambi@coli.uni-saarland.de

<sup>b</sup> Department of Experimental Psychology, Ghent University, 2 Henri Dunantlaan, B-9000,  
Ghent, Belgium; email: uschi.cop@ugent.be

<sup>c</sup> Department of Psychology, University of Edinburgh, 7 George Square, EH8 9JZ,  
Edinburgh, U.K.; email: martin.pickering@ed.ac.uk

<sup>1</sup>Address for correspondence (present address):

Chiara Gambi

Psycholinguistics Group

Department of Computational Linguistics and Phonetics

Building C 7.4

Campus Saarbruecken

Saarland University

Saarbruecken, 66123, Germany

[gambi@coli.uni-saarland.de](mailto:gambi@coli.uni-saarland.de)

Phone: +49 (0)681 302 6554

Fax: +49 (0)681 302 6561

## Abstract

We investigated whether speakers represent their partners' task in a joint naming paradigm. Two participants took turns in naming pictures; occasionally the (initial) picture was replaced by a different picture (target), signalling that they had to stop naming the initial picture. When the same participant had to name the target picture, he or she completed the name of the initial picture more often than when neither participant had to name the target picture. Crucially, when the other participant had to name the target picture, the first participant also completed the name of the initial picture more often than when neither participant named the target picture. However, the tendency to complete the initial name was weaker when the other participant had to name the target than when the same participant went on to name the target. We argue that speakers predict that their partner is about to respond using some, but not all, of the mechanisms they use when they prepare to speak.

*Keywords:* coordination, joint task, prediction, forward model, error repair

This is the authors' copy of the following paper:

<http://www.sciencedirect.com/science/article/pii/S0010945214003001>

to appear in *Cortex*.

## How do speakers coordinate?

### Evidence for prediction in a joint word-replacement task

#### **1. Introduction**

There is substantial evidence that observers predict actions (e.g., Kilner, Varga, Duval, Blakemore, & Sirigu, 2004; Ramnani & Miall, 2004; Flanagan & Johansson, 2003; Graf et al., 2007; see Wilson & Knoblich, 2005 for a review). For example, the readiness potential, which indexes the preparation of motor responses, is present from about 500 ms prior the observation of a predictable hand action (Kilner et al., 2004). Similarly, comprehenders often predict language (e.g., Altmann & Kamide, 1999; Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort, 2005; see Huettig, Rommers, & Meyer, 2011; Kutas, DeLong, & Smith, 2011; Pickering & Garrod, 2007; Van Petten & Luka, 2012 for reviews and discussion). For example, readers experience difficulty (i.e., enhanced N400) when the form of the indefinite article in English is not consistent with the initial phoneme of a highly expected noun (e.g., "an" when the expected noun begins with a consonant; DeLong, Urbach, & Kutas, 2005), indicating that phonological features of an upcoming word can be predicted.

But how do comprehenders compute such predictions? Researchers have proposed different mechanisms (Kutas, et al., 2011; Levy, 2008; Pickering & Garrod, 2007, 2013). In this paper, our aim is to answer one general question about the nature of such mechanisms, that is: To what extent are the mechanisms used for prediction related to the mechanisms used when preparing to speak? In other words, are the process of preparing to speak and the process of predicting whether another person is about to speak related to one another? If so, one would expect predictions to affect language production on-line. More precisely, if the

same mechanism is implicated concurrently in speech preparation and in predicting that another person is about to speak, then we would expect the latter process to affect the former.

There is some evidence that production processes might be involved in prediction during language comprehension. Federmeier, Kutas, and Schul (2010) reported that a late prefrontal positivity induced by plausible but unexpected nouns (which is thought to index error correction and/or prediction updating; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) is greatly reduced in older compared to younger adults. Importantly, the magnitude of this component in the older group correlated with production measures of verbal fluency (see also DeLong, Groppe, Urbach, & Kutas, 2012). Similarly, Mani and Huettig (2012) found that 2-years-olds with larger production (but not comprehension) vocabularies were more likely to predict upcoming referents (by looking at corresponding pictures) than their peers with smaller production vocabularies. These studies suggest that the ability or tendency to predict during language comprehension is correlated with language production abilities both in older adults and in children.

Pickering and Garrod (2013) proposed that prediction during language comprehension is subserved by the same mechanism that subserves feedforward control during language production, namely forward models (e.g., Wolpert, 1997). In their proposal, forward models map from production commands (communicative intentions) to the (production and comprehension) representations that will be retrieved as a consequence of executing those production commands. During language production, forward-model predictions are used for self-monitoring and learning. During comprehension, they are used in other-monitoring, and crucially to speed up and enhance understanding of the speaker's utterances (see Pickering & Garrod, 2014).

Recent MEG evidence suggests that covert language production (imagining to articulate or covert rehearsal in working memory) can selectively enhance early auditory

responses to syllables (Tian & Poeppel, 2013; Ylinen et al., 2014). In addition, motor activation occurs during speech perception, particularly during adverse conditions (D'Ausilio, Bufalari, Salmas, & Fadiga, 2012). Finally, activation in the right cerebellum correlates with adaptation to distorted speech in a perceptual task (Guediche, Holt, Laurent, Lim, & Fiez, 2014), while rTMS of the right cerebellum delays predictive eye-movements to upcoming linguistic referents (Lesage, Morgan, Olson, Meyer, & Miall, 2012). Importantly, the cerebellum has been implicated in the computation of forward models by several authors (e.g., Ito, 2008), and there is some evidence that the computation of motor-to-auditory mappings might be atypical in patients with cerebellar lesions (Knolle, Schröger, & Kotz, 2013).

In sum, there is converging evidence for the implication of production mechanisms in prediction of one's own and others' utterances. Specifically, the evidence reviewed above suggests that prediction could involve some form of internal simulation of a production process, and that it might be remarkably specific. In other words, comprehenders might simulate, using language production mechanisms, details of the linguistic content of another's utterance, for example associated with meaning (e.g., such as whether an upcoming referent is likely to be an edible object) or sound (e.g., whether an upcoming noun is likely to start with a consonant, or whether an upcoming vowel is likely to involve formant frequencies within a certain range).

However, at present, clear causal evidence for the implication of language production processes in content-specific prediction is limited to phonetics. Neurophysiological studies of syllable or pseudoword perception show that an articulation-related mechanism (i.e., activation of speech motor programs) is responsible for the effects of overt and covert language production on neural responses in auditory areas. But we do not yet know whether the same would hold for words and other meaningful units.

In addition, production and prediction could share general-purpose mechanisms (e.g., heightened attention; preparedness to respond), rather than language-specific mechanisms (i.e., processes involved in formulating utterances). While general-purpose mechanisms would not be able to support prediction of specific linguistic content (i.e., what somebody is about to say), they could in principle support prediction of whether another is about to speak (or, indeed, act in some other way). Such mechanisms could, for example, help speakers to predict whether another conversational participant is about to take the floor (Wenke et al., 2011), either by producing a linguistic utterance or by producing a non-verbal utterance (e.g., pointing gesture; Clark, 1996).

In this study, we tested the hypothesis that speakers predict whether another person is about to speak using mechanisms that are also implicated when they prepare to speak themselves, and investigated whether beliefs about another person's upcoming task can affect the way a speaker produces his or her own utterance. To this aim, we devised a joint language production task that requires participants to take turns in speaking. Joint tasks have been used to study similar issues in the domain of action, as we briefly discuss below before returning to language.

### **1.1 Joint Tasks in the Action Domain**

In joint task paradigms (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006; Sebanz & Knoblich, 2009), participants are tested in pairs and are assigned complementary tasks (i.e., they each perform half of the task that would be performed by a single participant in solo task paradigms; see below). For example, in one study (Knoblich & Jordan, 2003), pairs of participants attempted to keep a circle aligned with a moving dot on a computer screen. In each pair, one participant could accelerate the tracker only to the right, while the other could accelerate it only to the left. Performance in the joint task is usually

compared to performance in a solo version of the same task. In this study, in the solo version an individual participant could control the tracker's velocity in both directions (using two hands).

Interestingly, Knoblich and Jordan (2003) found that performance in the joint task improved over time, and eventually became as good as in the solo task. In the tracking task, accurate timing is essential for good performance. For example, the right participant should avoid accelerating when the left participant is accelerating in the opposite direction. In order to avoid overlap, the right participant needs to predict the left participant's actions, and vice versa. This study therefore suggests that, under some conditions at least (e.g., given sufficient time to adapt to their partner), people can predict each other's actions, and, furthermore, that such predictions can affect action planning online in a way that is similar to how predictions of the effects of one's own action can affect planning.

The rationale behind the comparison between joint and solo versions of the same task is as follows: If performance of a participant in a pair (or indeed, performance of the pair as a whole) is similar to performance of an isolated participant, then it suggests that similar mechanisms underlie performance in both situations. It has been argued that such similarity in performance can be explained, in part, by the assumption that self- and other-generated actions share the same representational format (though additional mechanisms might be necessary; see Sebanz, Bekkering, Knoblich, 2006). Interestingly, it has also been suggested that one mechanism operating in joint tasks could involve keeping track of whose turn it is to respond (Wenke et al., 2011).

In this study we apply an analogous rationale to the domain of language and compare performance in a joint language task to performance in a solo language task to investigate whether predicting that another will speak makes use of the same mechanisms involved in speech preparation. Importantly, this is the first study to investigate whether the coordination



of two successive utterances between two speakers can be supported by the same mechanisms that support the coordination of two successive utterances produced by the same speaker.

## 1.2 A Joint Language Production Task

In our joint language production task, two speakers, *A* and *B*, sit next to one another in front of the same computer screen. Speaker *A* produces an *initial* utterance (a picture's name), and then speaker *B* produces a second *target* utterance (another picture's name, which is unrelated to the first picture's name). More precisely, when the picture *A* is instructed to name is replaced by a new picture, speaker *A* has to stop producing her utterance as quickly as possible; then speaker *B* names the new picture (see below). We call this the joint word-replacement task. We ask whether and how the way speaker *A* produces the initial utterance is affected by the fact that speaker *B* will later speak. We are interested in speaker *A*'s (rather than *B*'s) utterance because it occurs first. Therefore, any effects of *B*'s task on *A*'s utterance would be due to *A*'s prediction of what *B* is about to do (and not to *A*'s comprehension of *B*'s utterance).

If speaker *A* indeed predicts that *B* will speak, the mechanism(s) she uses to compute this prediction could stand in one of three relationships with respect to the mechanism(s) she uses to prepare her own utterance. One possibility is that prediction mechanisms are entirely independent from production mechanisms. For example, predictions could be computed using an inference-based mechanism. Note that speaker *A* does not comprehend any part of *B*'s utterance before she has finished her own utterance (as *B* starts speaking only after *A* has stopped). But it is possible that *A* predicts that *B* is about to respond by inferring what *B* is most likely to do given the instructions *A* received about the task, perhaps using mechanisms involved in elaborative (i.e., predictive) inferences (e.g., McKoon & Ratcliff, 1986).

If this account is correct, predicting that another speaker is about to speak should have no direct and immediate effect on production processes. Say that speaker *A* infers that *B* is about to speak. Even if *A* constructs this inference very quickly (within the space of planning and uttering a single word), there is no reason to expect that doing so should affect *A*'s production of her own utterance. Therefore, we term this the *separate mechanisms* account.

A second possibility is that the mechanisms that speaker *A* uses to predict *B*'s act of responding are precisely the mechanisms that she uses to prepare her own utterance. In other words, predicting that *B* is about to speak would entail the same processes on the part of speaker *A* that she would use to prepare to speak herself. We term this the *shared mechanisms* account. If this account is correct, predicting that another is about to speak should have the same effect as preparing to speak, because speaker *A* would automatically activate all the production processes leading to the formulation of the target picture name even though *B* has to name this picture. An analogous suggestion has been made in the joint action literature to explain why in joint tasks people appear to represent their partner's responses as if they were their own (Knoblich et al., 2011).

A third possibility is that predictions of others' utterances are computed using some of the mechanisms used when producing utterances, but not all. We term this the *overlapping mechanisms* account. If this account is correct, predicting that another is about to speak should have some effect on production processes, but this effect might be different from the effect of preparing to speak oneself.

In order to distinguish among these three possibilities, we tested two more versions of our word-replacement task in addition to the joint version described above. In the no-replacement version of the task, speaker *A* names the initial picture but speaker *B* (who is present) does not name the target picture. Therefore, *B* remains silent and there is simply no response to predict. In the solo word-replacement task, instead, *A* names the initial picture

and then names the target picture as well. (*B* is still present and performs the same solo task on different trials.) Therefore, *A* needs to retrieve the second picture name and will go through all the stages of language production (from concept selection to articulation), as assumed by all theories of single word production (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999).

If prediction mechanisms are completely independent of production mechanisms, the joint word-replacement task should be equivalent to the no-replacement task: Even if *A* predicts that *B* is about to speak in the joint version, *A*'s predictions should not affect the way *A* produces her utterance, given that speakers are not given any explicit instruction to coordinate their utterances with their partner's utterances; in other words, predicting that another is about to speak should have a comparable effect on production as predicting no response at all. On the contrary, if the shared mechanisms account is correct, then the joint version of the word-replacement task should be equivalent to the solo word-replacement task, as *A* would predict *B*'s response in the joint task using the same mechanisms that she uses to prepare the target utterance in the solo task.

Finally, if the overlapping mechanisms account is correct, the joint word-replacement task will not be equivalent to the no-replacement task, because in the joint task *A* predicts that *B* is about to speak using some mechanisms that can affect production. However, the joint version need not be equivalent to the solo version, because in the joint task *A* predicts that *B* is about to speak using some mechanisms that are used during language production, but not the full range of language production mechanisms.

Language production is of course a complex process, which involves both mechanisms that are specific to formulating an utterance (Levelt, 1989) and mechanisms that are shared with other cognitive activities (e.g., attention, memory, motor preparation). Accordingly, there are two versions of the overlapping mechanisms account. One version

claims that the mechanisms that are shared between production and prediction are specifically linguistic (i.e., some of the mechanisms of language production are used to predict that another is about to speak). The other version claims that the shared mechanisms are general-purpose mechanisms, instead.

We return to this issue in the Discussion. Here, we note that our study was not designed to distinguish between alternative versions of the overlapping mechanisms account. Rather, our aim was to test among the separate mechanisms, the shared mechanisms, and the overlapping mechanisms account. Importantly, the solo word-replacement and the no-replacement tasks have been used by Hartsuiker, Catchpole, De Jong, and Pickering (2008), so we can use their findings to formulate specific hypotheses about the joint word-replacement task. Hartsuiker et al. were interested in how speakers coordinate the planning and articulation of two utterances in speech repairs. We briefly review this literature below to consider which factors might affect performance in the solo version of the task, and then ask whether similar factors would affect performance in our joint task.

### 1.3 Coordinating Stopping and Resuming Speech

In instances of self-repair, the speaker coordinates the planning and articulation of two utterances: the initial utterance and the replacement. Thus in “Left – er – right in front of me” (Levelt, 1989, p. 484), the initial word (*left*) is completed, and then the replacement (*right*) follows after an editing expression (*er*). But sometimes the initial word is not completed, as in “To the left side of the purple disk is a *v* – a horizontal line” (most likely, the intended word was *vertical*; Levelt, 1989, p. 474). These examples illustrate that there is variability as to *where* speakers stop (between words or within words) when they detect an inappropriate word and correct themselves (Levelt, 1983).

This observation has sparked considerable theoretical (Hartsuiker & Kolk, 2001; Levelt, 1989; Nooteboom, 1980) and empirical investigation, both in the form of observational studies (Blakmer & Mitton, 1991; Seyfeddinipur, Kita, & Indefrey, 2008) and experiments (Hartsuiker, et al., 2008; Hartsuiker, Pickering, & De Jong, 2005; Tydgat, Stevens, Hartsuiker, & Pickering, 2011; cf. Van Wijk & Kempen, 1987). In these experiments, self-repairs are induced by asking participants to describe an initial picture, which is then replaced by a target picture. This sometimes causes participants to reformulate their utterances.

Note that such experiments do not investigate the situation in which the speaker detects an internally generated error. Specifically, the need to reformulate is caused by a change in the environment (see Tydgat et al., 2011, p. 360). However, this feature of the task is useful for our purposes, as it makes the solo version of the task directly comparable to the joint version of the task.

According to the account proposed by Hartsuiker and Kolk (2001), and modified by Hartsuiker et al. (2008) and Tydgat et al. (2011), the speaker simultaneously initiates two processes when executing a self-repair: the process of stopping articulation and the process of planning the replacement. These processes proceed in parallel and share a limited pool of resources. Therefore, the process of planning the replacement competes with the process of stopping the initial word. In other words, the speaker uses production mechanisms to begin preparing the target word while also trying to stop the planning and articulation of the initial word, therefore incurring interference. This hypothesis is supported by two findings.

First, Hartsuiker et al. (2008) found that speakers complete initial words more often when they subsequently had to produce a replacement (53.9%) than when they simply had to stop speaking (21.5%; see also Tydgat et al., 2011). Thus, speakers find it harder to stop the initial word when they also need to start preparing a replacement than when they do not. This

could reflect competition between a “go” and a “stop” signal, in line with the account proposed by Hartsuiker et al.<sup>1</sup> Alternatively, speakers might strategically evaluate whether it is better to interrupt the initial word as quickly as possible or to continue with articulation (Tydgat et al., 2011; cf. Seyfeddinipur et al., 2008, when self-repairing internally generated errors); by continuing to articulate previously planned material, speakers could in fact re-allocate resources from stopping to replacement planning, and ensure that the replacement is ready in reasonable time. Second, Hartsuiker et al. (2008; Experiment 1) found that the spoken duration of the initial word was longer when planning the replacement was made more difficult by degrading the target picture.

To sum up, replacing a word with another involves the coordination of two processes that compete for resources: 1. stopping the initial word; 2. planning the replacement. Note that we do not assume that processes are specific to speech, nor that the stopping process and the re-planning process need to belong to the same domain in order for interference to occur (as long as they have access to a common pool of resources). Stopping and resuming speech could indeed rely on a general-purpose monitoring system (see Riès, Janssen, Dufau, Alario, & Burle, 2011).

More importantly, in the solo version of the word-replacement task, one speaker carries out both the stopping and the replacement planning processes. In the no-replacement task, instead, the speaker stops the initial word but does not plan the replacement (so only one process is involved). Finally, in the joint version, the two processes are distributed between two speakers: Speaker *A* plans the initial word and stops, and speaker *B* plans the replacement. The question we ask in this study is whether *A* uses some processes involved in

---

<sup>1</sup> Note that Hartsuiker et al. (2008) did not comment on this aspect of their results (and did not test for it statistically).

planning the replacement to predict that *B* is about to speak in the joint version of the task. The next section describes the experimental conditions and presents the expected findings according to the separate mechanisms, the shared mechanisms, and the overlapping mechanisms account.

#### **1.4 An Experimental Comparison of the Accounts**

In three conditions, a pair of participants viewed a picture that appeared on a shared screen, and we cued one or the other participant to name that picture. On a small proportion (9%) of trials, the initial picture changed into a target picture (as in Hartsuiker et al., 2008). When the change occurred, the participant was instructed to stop naming the initial picture as quickly as possible.

In all conditions, the cued participant varied across trials. Instructions about the target picture depended on the condition to which the participant was assigned. In the SELF condition (solo task), the cued participant also named the target picture. This condition therefore followed Hartsuiker et al.'s (2008) Experiment 1, except that it involved two (co-present) participants. In the OTHER condition (joint task), the other (non-cued) participant named the target picture. In the NO condition (no-replacement task), neither participant named the target picture. This last condition therefore followed Hartsuiker et al.'s Experiment 2, except that it again involved two participants. Following the results of Hartsuiker et al., we hypothesized that participants in the SELF condition would complete the initial word more often than participants in the NO condition. This finding would confirm that participants in the SELF condition were planning the target picture name before stopping the initial name and that these processes competed for resources.

Note that the presence of another person can affect individual performance in complex ways, sometimes yielding facilitation, sometimes interference (e.g., Klauer,

Herfordt, & Voss, 2008). Indeed, the presence of another person serves as a retrieval cue for words that have been uttered by that person and facilitates picture naming (Horton, 2007). So it was important to investigate whether Hartsuiker et al.'s (2008) results would be replicated in the presence of another person. To further ensure comparability between our results and theirs, in all conditions the target picture was either intact or degraded (with 50% of its contours removed). Based on Hartsuiker et al.'s Experiment 1, we expected participants in the SELF condition to stop naming the initial picture later when the target picture was degraded versus intact (i.e., a *degradation effect*). More resources are needed to retrieve the name of a degraded picture because the associated concept is more difficult to identify; therefore, the process of planning the target word should interfere more with the process of stopping the initial word when the target picture is degraded than when it is intact. In addition, based on Hartsuiker et al.'s Experiment 2, we expected no degradation effect in the NO condition.

Consider now the novel OTHER condition. Let us assume that, at some point during the process of stopping the initial word, speaker *A* predicts that speaker *B* is about to speak. If prediction mechanisms are completely independent of production mechanisms (as the separate mechanisms account assumes), speaker *A*'s prediction will not affect her ability to stop producing the initial word. Therefore, *A* should find stopping the initial word no harder in the OTHER condition than in the NO condition. More specifically, she should be no more likely to complete the initial word in the OTHER than in the NO condition.

But if, on the contrary, *A* predicts that *B* is about to speak using the same mechanisms she would use when she prepares to speak herself (as the shared mechanisms account assumes), then *A*'s prediction should affect her ability to stop producing the initial word. More precisely, *A*'s prediction should affect her ability to stop producing the initial word in the same way as the process of planning the target word would affect her ability to stop (i.e.,



as in the SELF condition). Therefore, *A* should be more likely to complete the initial word in the OTHER than in the NO condition and, moreover, *A* should be as likely to complete the initial word in the OTHER as in the SELF condition.

Finally, if *A* predicts that *B* is about to speak using some of the mechanisms that are used in production, but not all of them (as the overlapping mechanisms account assumes), then *A*'s prediction should still affect her ability to stop producing the initial word. Crucially, however, this effect need not be the same as in the SELF condition. One possibility is that *A* might find it less hard to stop in the OTHER than in the SELF condition (while still finding it harder than in the NO condition, where production mechanisms are not used at all). Table 1 summarizes the differences between the accounts.

INSERT TABLE 1 ABOUT HERE

## **2. Method**

### **2.1 Participants**

Ninety-six students from the University of Edinburgh participated in the experiment. They were either paid £6 or received course credit in return for participation. All were native English speakers and reported no language impairment. Participants were matched to form 48 pairs, which were then randomly assigned to each of the three conditions. Thus, we tested 16 pairs of participants in each condition. Most participants did not know their partners beforehand. The study was approved by the Ethics Committee of the Department of Psychology of the University of Edinburgh.

### **2.2 Materials**

The materials were simple black and white line drawings. There were 32 target pictures, each of which appeared in an intact and a degraded format. These were the target

pictures used by Hartsuiker et al. (2008), derived from a set originally developed by Meyer, Sleiderink, and Levelt (1998). To create the degraded versions, Meyer et al. deleted “50% of the black pixels, in regions where they could be reconstructed by straight or smoothly curved lines” (p. 27). There were 32 initial pictures, also taken from Hartsuiker et al., and 128 filler pictures from Snodgrass and Vanderwart (1980). Each of the 64 experimental items constituted a unique combination of an initial picture and a target picture. The pictures were combined in such a way that every target picture occurred after 2 initial pictures and every initial picture preceded 2 target pictures (e.g., *glasses-mouse*, *glasses-wall*, *orange-mouse*, *orange-wall*). In each item, the names for the initial and target pictures had different initial phonemes and unrelated meanings (see Appendix A for a complete list of the experimental items). Initial pictures and filler pictures were presented inside a colored frame (green or red) in order to cue one participant to name that picture (see Procedure). Target pictures were presented without a frame (the instructions made clear who was to name a given target picture; see Procedure). Initial and target pictures were used both on change (experimental) and no-change (filler) trials, whereas filler pictures were used only on no-change trials.

### 2.3 Design

Degradation (intact vs. degraded) was varied within participants and items. Condition (SELF, OTHER, NO) was varied between participants but within items. We first created four lists containing the 64 experimental items (change trials). Every initial picture and target picture occurred twice in each list of change trials. Each target picture appeared once degraded and once intact. The degraded version of a target picture was paired with an initial picture of one color and the intact version was paired with an initial picture of the other color; also, each initial picture occurred once in each color. This meant that each participant in a pair named each initial picture and each target picture only once on change trials (though, of course, they saw each initial and each target picture twice). In addition, we divided the

experiment into two blocks and each initial picture and each target picture appeared once in each block. For each experimental item, every combination of color-assignment and target degradation (red initial – degraded target, red initial – intact target, green initial – degraded target, green initial – intact target) occurred once across lists. For each of the 4 lists, we derived 4 random orders, with the constraint that each block appeared first in half these orders.

We also constructed 2 lists of no-change trials. Each contained 640 items: the 32 target pictures twice (once degraded, once intact) in isolation; the 32 initial pictures twice in isolation; and the 128 fillers, four times each. The two lists were constructed so that the target pictures and initial pictures that were presented in one color in the first list were the other color in the second list. They were also split into two blocks, with repetitions of the same picture being equally distributed between blocks. To create running lists, one no-change trial list was combined with one change trial list. The pairing was done in such a way that target pictures in the change trial list had the opposite degradation relative to their instances (within the same block) in the no-change trial list, and were always named by the other participant. Each change trial was separated by at least three no-change trials. The same 16 running lists were presented in each of the three between-participants conditions.

## **2.4 Procedure**

The experiment was controlled using E-Prime (Version 2.0). First the participants were introduced to each other and told that they were going to do a task together. They were then familiarized with the materials in individual booths. They were shown the 192 pictures (32 initial pictures, 32 target pictures, 128 fillers) with the corresponding names, and were instructed to read the names out loud to aid memory. Next, the two participants were seated in front of the same computer screen. Half of the pairs were instructed that the pictures in the green frame were to be named by the person on the left, and the pictures in the red frame

were to be named by the person on the right. The other half of the pairs were instructed that the pictures in the red frame were to be named by the person on the left, and the pictures in the green frame were to be named by the person on the right. They were told to use the names that they had learned during the naming phase.

The instructions about change trials depended on condition. For pairs in the SELF condition, cued participants were instructed to stop naming the initial picture and name the interrupting picture as fast as possible. Therefore, in the SELF condition the participant who responded to the initial picture also responded to the target picture on the same trial. Cued participants in the OTHER condition were also instructed to stop naming the initial picture as fast as possible, but this time the other participant had to name the interrupting picture. Therefore, in the OTHER condition one participant responded to the initial picture and the other responded to the target picture on the same trial. The color of the initial picture frame indicated who was to perform which task on any given change trial. Both participants performed each task equally often in each block, while they took turns according to a randomized sequence. Finally, in the NO condition, cued participants were again instructed to stop naming the initial picture as fast as possible, but they were told to ignore the interrupting picture. Therefore, in the NO condition none of the participants responded to the target picture on change trials (see Fig. 1).

INSERT FIGURE 1 ABOUT HERE

Before starting the experiment, the participants completed 8 practice trials. These were 5 no-change trials and 3 change trials on which filler pictures were used instead of experimental pictures. After the practice, the instructions were summarized again and the participants were warned that some of the pictures would consist of dashed lines.

All trials started with a fixation cross which remained on the screen for 2500ms. On no-change trials an initial picture (with a colored frame) then appeared for 500ms. On change trials the initial picture appeared for 300ms and was then replaced by a target picture (without a colored frame) that appeared for 500ms. The inter-trial interval was 3300ms after a change trial and 3000ms after a no-change trial. Participants spoke into head-mounted microphones and their responses were digitally recorded on two separate channels. For each change trial, two audio files were generated (and automatically stored), one time-locked to initial-picture onset, the other time-locked to target-picture onset. An experimental session lasted approximately 45 minutes.

## 2.5 Scoring

Only change trials are relevant for our hypotheses, so only the audio files recorded during these trials were analyzed. Data from 7 pairs (3 in the SELF condition, 2 in the OTHER condition, 2 in the NO condition) had poor audio quality, and so background noise was reduced by batch processing their change-trial files, using Adobe Audition (Version 1.05). Responses that were still inaudible or could not be categorized were excluded from further analyses; if there were more than 10 such trials for a single pair, the whole set of data for that pair was discarded. This resulted in the loss of 1.8% of the data in the SELF condition, 1.4% in the OTHER condition, and 3.5% in the NO condition, in which one pair was discarded.

The remaining responses were annotated off-line (half by the first, half by the second author). We first noted errors and disfluencies (e.g., *um*, repetitions) in producing the initial or target name (in SELF and OTHER); see the Results section for percentages. For two target pictures (*mouth*, *steps*), participants responded with an alternative name (*lips*, *stairs*) on at least 39% of the trials. As these were clearly acceptable responses, we included these trials in

the analyses. All other naming errors were coded as such and the corresponding trials were discarded. Correct and fluent initial responses were divided into three response types: completed initial (e.g., *apple chair*), interrupted initial (e.g., *ap- chair*), and skipped initial (e.g., *chair*).

Second, three time-points were manually annotated on the audio files using the phonetic analysis software Praat (Boersma & Weenink, 2010): the onset of the initial name, the offset of the initial name, and the onset of the target name (in the SELF and OTHER conditions). We used these time-points to determine the following time measures: Initial Onset (onset of initial name relative to onset of the initial picture); Initial Duration (onset of initial name to offset of initial name); Target Onset (onset of target name relative to onset of the target picture, in the SELF and OTHER conditions); Interval (offset of initial name to onset of target name, in the SELF and OTHER conditions). The primary measure of interest was Initial Duration. However, we also analyzed the other time measures, in part to determine whether our results were consistent with Hartsuiker et al. (2008). Summary tables and a brief description of these results can be found in Appendix B. Additional data exclusion and trimming criteria for the time measures are reported in the Results section (for Initial Duration) and in Appendix B (for the other measures).

## **2.6 Data Analysis**

The data were analyzed using Generalized Linear mixed-effects models (Bates, Maechler, & Dai, 2008; Baayen, Davidson, & Bates, 2008) in R (Version 2.7.2). For the response type data, we used a logistic link function (Jaeger, 2008) and conducted a binomial analysis (comparing the likelihood of a completed response against the likelihood of observing any of the two other kinds of responses; i.e., an interrupted or a skipped response). For Initial Duration, we only included completed initial responses in the analysis. This was motivated by the fact that there were no specific predictions for the factor Response Type. In

addition, completed responses were more evenly distributed than interrupted responses (see Table 2) and we hoped, in this way, to minimize issues related to the imbalance in the proportion of response types across conditions (see below). Consequently, the predictors of interest for the analysis of Initial Duration were only Condition and Degradation. See Appendix B for details of the analyses of the other time measures.

In all analyses, we started by fitting the complete model; we then removed predictors that were not significant from the model, using a backward stepwise procedure, and stopped whenever removing a predictor caused a significant loss of fit (assessed using a log-likelihood ratio test). We report coefficients, standard errors, and Wald's t-tests from the complete model together with results of the likelihood ratio test for each predictor (Barr, 2008; Quené & van den Bergh, 2008). Regarding random effects, we started with the full random effect structure, including random slopes (for all factors and their interaction) and random intercepts for both subjects and items (defined as a combination of initial and target picture). Given that random slopes are only appropriate for within-subjects and within-items factors, we included by-subjects random slopes for Degradation and by-items random slopes for Degradation, Condition, and their interaction. If the model with full random effects specification did not converge, we simplified it by removing the higher-order term (interaction of Condition and Degradation). We then tested whether specific random effects significantly contributed to model fit using likelihood ratio tests. We report estimates of the variances and covariances of all random effects that passed the test (with an alpha-value of .1 instead of .05 to account for the conservativity of these tests).

We used sum coding for our predictors, both in the response type analyses and in the analyses of Initial Duration. For the analyses of the other time measures, we used contrast (Helmert) coding, so that the coefficients associated with the factor Condition could be more easily interpreted (see Appendix B for further details). Because Response Type was not under

experimental control and was in fact affected by Condition (see below), the number of observations per cell varied widely, leading to a highly imbalanced design for the time analyses. This means that in order to have weighted estimates for the fixed effects, it is necessary to weight the contrasts by the observed cell counts. We therefore used weighted coding (Cohen, Cohen, West, & Aiken, 2003; Serlin & Levin, 1985; West, Aiken, & Krull, 1996) for all the predictors entered in the analyses of the time measures (see Appendix B for an example).

### **3. Results**

#### **3.1 Response Type Data**

As stated in the Scoring section, for the analyses of response type we excluded the trials where the initial picture was not named correctly or the initial name contained hesitations or repetitions (5.1% in the SELF condition, 5.8% in the OTHER condition, 4.7% in the NO condition). This left us with 963 data points in the SELF condition, 962 in the OTHER condition, and 914 in the NO condition. The percentages of Completed, Interrupted, and Skipped initial responses in each condition are reported in Table 2, separately for degraded versus intact trials.

INSERT TABLE 2 ABOUT HERE

The best-fitting model included only Condition as a predictor, whereas Degradation had no effect on the proportion of completed responses, nor did the interaction. No random slopes were justified, so only random intercepts were retained (see Table 3). Participants



completed the initial name more often in the OTHER condition (63.2%) than in the NO condition (46.5%). They also completed the initial name more often in the SELF condition (83.4%) than in the OTHER condition. In order to test the hypotheses of the three accounts laid out in the Introduction, we then set the OTHER condition as the reference level, and we defined two contrasts, one comparing the mean of the OTHER condition to the mean of the NO condition (Condition1), the other comparing the mean of the OTHER condition to the mean of the SELF condition (Condition2). The first contrast therefore tests whether speakers complete the initial word more often in the OTHER than in the NO condition, which would be compatible with both the shared mechanisms and the overlapping mechanisms account (but not with the separate mechanisms account). The second contrast tests whether speakers complete the initial word more in the SELF than in the OTHER condition, which would be compatible with the overlapping mechanisms account but not with the shared mechanisms account. Importantly, both contrasts were associated with estimates significantly different from zero (Condition1:  $B = -0.99$ ,  $SE = .46$ ,  $z = -2.16$ ,  $p < .05$ ; Condition2:  $B = 1.57$ ,  $SE = .47$ ,  $z = 3.36$ ,  $p < .001$ ) in a model that included only Condition amongst the fixed effects. Overall, these results are compatible with the overlapping mechanisms account, but not with the shared mechanisms account or the separate mechanisms account.

INSERT TABLE 3 ABOUT HERE

### 3.2 Initial Duration

For the analyses of Initial Duration, we removed all trials that were more than 2.5 SD from the grand mean or more than 3 SD from the by-subject mean (2.5% in SELF, 2.0% in OTHER, 1.7% in NO). As stated above, we limited our analyses to completed initial responses. Apart from this, we conducted the same analyses as for the response type data. In

addition, we conducted separate analyses for the three conditions in order to compare our results directly to Hartsuiker et al.'s (2008) findings.

INSERT TABLE 4 ABOUT HERE

Participants took 12 ms longer to stop before they named degraded than before they named intact targets in the SELF condition (see Table 4, completed responses). The inclusion of by-item random slopes for the factor Condition significantly improved fit (see Table 5). The main effect of Degradation marginally improved fit ( $p=.09$ ), as did the interaction of Degradation and Condition ( $p=.07$ ); the main effect of Condition was not significant. When we fitted separate models to the three conditions (Table 6), we found a degradation effect in the SELF condition ( $p<.01$ ) but not in the OTHER or NO conditions (both  $t$ 's  $< 1$ ).

INSERT TABLE 5 ABOUT HERE

INSERT TABLE 6 ABOUT HERE

#### **4. Discussion**

We investigated whether participants in a joint language production task predict that their partner will speak using language production mechanisms and whether such prediction affects production of their own utterance. To do so, we compared a solo word-replacement task (the SELF condition), a joint word-replacement task (the OTHER condition), and a no-replacement task (the NO condition).

We found that participants completed their initial utterance more often in the OTHER condition (63.2%) than the NO condition (46.5%), but less often in the OTHER condition than the SELF condition (84.3%). Therefore, we replicated Hartsuiker et al.'s (2008) findings in a two-person setting, as participants completed the initial word more often when they later named the target word than when they did not name the target word. The tendency to complete the initial word was greater in our SELF condition than in their Experiment 1 (53.9%), and similarly larger in our NO condition than in their Experiment 2 (21.5%), perhaps because the lower percentage of change trials in our study (9%) than in theirs (12.5%) made the task of stopping overall harder for our participants. In addition, in the SELF condition, we replicated the effect of Degradation in their Experiment 1 on the duration of the initial word. Similarly, in the NO condition, we found no effect of Degradation on the duration of the initial word, as in their Experiment 2.<sup>2</sup>

Participants tended to complete the initial word more often in the OTHER than in the NO condition. This suggests that they predicted that their partner was about to speak and that this prediction interfered with the process of stopping speech. This finding is consistent with the claim that predicting that another speaker is about to speak relies on some of the same mechanisms used during production, and is therefore not consistent with the separate mechanisms account. In addition, participants tended to complete the initial word less often in the OTHER condition than in the SELF condition. This suggests that they did not activate production mechanisms to the same extent in the OTHER as in the SELF condition, and

---

<sup>2</sup> Note that Hartsuiker et al. (2008) analyzed both completed and interrupted responses, whereas we analyzed only completed responses. They did so because the proportions of completed and interrupted responses were much less unbalanced in their experiments than in our experiment.

hence the study is not consistent with the shared mechanisms account. In sum, this set of findings is compatible only with the overlapping mechanisms account. We discuss this account in detail in Section 4.1.

The finding that inhibiting a response (i.e., stopping on change trials) is harder for speakers when they know that their partner is about to respond (in the OTHER condition), than when they know their partner is not about to respond (in the NO condition) is consistent with neuroscientific studies of joint action. In particular, several ERP studies reported increased response inhibition demands on NO-GO trials in joint tasks compared to individual go/no-go tasks. This suggests that participants in joint tasks represent their partner's actions on NO-GO trials, and need to apply a higher level of inhibition (indexed by an enhanced no-go P3 component) to avoid responding overtly when it is their partner's turn to respond (Sebanz, Knoblich, Prinz, & Washer, 2006; Tsai, Kuo, Jing, Hung, Tzeng, 2006; Tsai, Kuo, Hung, & Tzeng, 2008). In addition, in one fMRI study (Sebanz, Rebbechi, Knoblich, Prinz, & Fritz, 2007) NO-GO trials in the joint condition (compared to an individual condition in which the partner was present but not active) showed increased activity in the SMA (Supplementary Motor Area), which is implicated in the execution of motor responses (see e.g., Mostofsky & Simmonds, 2008).

Finally, de Bruijn, Miedl, and Bekkering (2008) tested participants in a competitive speeded go/no-go task. Participants who performed more poorly (i.e., were on average slower than their partner) showed reduced no-go P3 amplitudes (and hence, lower inhibition) when their partner was responding compared to when their partner was not responding (so both actors had to inhibit a response), suggesting that they could not help but represent their partner's task and that this impaired their performance.

Note that all these studies used manual responses. We are not aware of comparable imagining or electrophysiological evidence for tasks involving verbal responses. Future studies could investigate the neural correlates of the decreased likelihood of stopping word-internally in the OTHER compared to the NO condition.

#### **4.1 Predicting that You Are about to Speak**

The overlapping mechanisms account states that prediction uses some, but not all of the mechanisms used during production. It therefore raises the question: What is the precise nature of prediction mechanisms? We know that they are used in language production, but what kind of mechanisms are they? As mentioned in the Introduction, it is possible that they are general-purpose mechanisms, like those implicated in the allocation of attention or in preparing a (not necessarily verbal) response. It is also possible that such mechanisms are specific to the process of producing language (as opposed to, for example, producing a non-linguistic action).

This study cannot adjudicate between these two versions of the overlapping mechanisms account. We showed that speakers find it harder to stop an utterance when they know their partner will produce another utterance than when they know their partner will produce no response. However, we do not know whether speakers would also find it harder to stop if they knew their partner were about to act in some other way (e.g., pressing one button if the target picture is degraded, and another button if the picture is intact).

Indeed, speakers appear to coordinate the production of two successive utterances in a similar way to the production of one utterance and of a manual response. Speakers who were asked to name a left object and a right object in close succession (i.e., without pausing) shifted their gaze to the right object later with respect to the onset of speech when the left object had a long (trisyllabic) name than a short (monosyllabic) name, possibly because long

names are planned more incrementally (Meyer, Belke, Häcker, & Mortensen, 2007, Experiment 1; see also Griffin, 2003). The same pattern was observed when speakers were asked to name the left object and then press a button to categorize a symbol that appeared on the right side of the screen (Meyer, Belke, Häcker, & Mortensen, 2007, Experiment 4).

Nevertheless, ours is the first study to compare the coordination of two successive utterances within and between speakers, and to show that the way in which speakers produce their utterances can be affected by whether they predict their partner will soon act. It suggests that between-speaker coordination makes use of some mechanisms that are also involved in preparing to speak. Below, we discuss how our findings relate to other evidence for the implication of language production mechanisms in prediction. Particularly, we focus on two related proposals that stress the language-specific nature of prediction mechanisms and consider how such proposals could account for our findings. Both views assume that the mechanisms common to language production and language prediction are domain specific, but they differ in the details of the mechanisms involved.

The first proposal is that predicting that another will speak (or, indeed, what another will say) entails activation of linguistic representations within the language production system. If *A* predicts that *B* is going to speak, she does this by going through the stages of language production (e.g., accessing semantics, syntax, phonology) that she goes through when she prepares to speak herself. Importantly, the finding that our participants completed their utterance less when they predicted that their partner was about to speak than when they were about to speak indicates that the production system was only *partly* activated when *A* predicted that *B* was about to speak. Accordingly, at some point during the process of language production, *A* might inhibit her production system (so that she does not actually speak).

There is evidence that language production mechanisms are activated during language comprehension. Listening to speech modulates the excitability of muscles involved in articulation (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Watkins, Strafella, & Paus, 2003; Yuen, Davis, Brysbaert, & Rastle, 2010). Moreover, D'Ausilio, Jarmolowska, Busan, Bufalari, and Craighero (2011) repeatedly exposed participants to a pseudoword (e.g., *birro*) and used TMS to reveal immediate appropriate articulatory activation (associated with *rr*) when they heard the first part of the same item (*bi*, when coarticulated with *rro*) compared to when they heard the first part of a different item (*bi*, when coarticulated with *ffo*). Similarly, when observing a signed utterance that ended in a semantically unexpected sign, German signers showed an enhanced N400 effect whose onset began before the onset of the sign itself, during the transition from the previous sign (Hosemann, Herrmann, Steinbach, Bornkessel-Schlesewsky, & Schlesewsky, 2013). This suggests that activation of the language production system might be involved in prediction, so that listeners in D'Ausilio et al.'s study activated the specific articulators involved in the production of the expected sound, and observers in Hosemann et al.'s study activated details of the kinematics of the expected sign.

However, it is not known whether activation always occurs at all stages of language processing, with inhibition suppressing only overt production (as findings such as D'Ausilio et al., 2011 suggest), or whether inhibition can occur at any stage. This question is reminiscent of a discussion concerning the nature of inner speech, where some accounts posit fully-specified sub-phonemic features (Corley, Brocklehurst, & Moat, 2011), whereas others maintain a more abstractionist view in which inner speech is specified only up to the phonological level (Oppenheim & Dell, 2010). But it is also possible that inhibition can occur at an even earlier stage (e.g., before or during lexical selection).

The second proposal is that predicting that another will speak (or, indeed, what another will say) involves the computation of forward models. Pickering and Garrod (2013) proposed that speakers send a production command (communicative intention) to two different systems: the production implementer and the forward production model. The output of the production implementer (which contains the mechanisms contained in all models of language production) is an actual utterance, together with its associated set of structured linguistic representations, encoding semantics, syntax, and phonology. The output of the forward model, instead, is a prediction of some aspects of the representation of an utterance, possibly including a prediction of some aspects of the semantics, syntax, and phonology of the utterance. For example, upon recognizing a chair, speakers might predict that a concrete noun beginning with the phoneme /tʃ/ is about to be produced. Importantly, forward-model predictions are normally ready before the representations computed by the production implementer. Therefore, they can be used for the online control of language production processes.

For example, imagine a sports commentator reporting live on a soccer match. She is providing some statistics about previous matches, when one of the players suddenly performs an amazing pass that could lead to a goal. The commentator then issues a stop signal to her articulators (depending on various factors; e.g., how much she values fluency; cf. Seyfeddinipur et al., 2008), and starts retrieving the player's name from memory (using the production implementer). This process requires resources, and therefore interferes with the process of stopping speech.

In addition, and before retrieving the player's name, the commentator predicts that she will soon produce an utterance (using a forward model). Of course, the commentator has had to stop and reformulate before. She might have learned that it is difficult to stop speech while attempting to formulate a new utterance. She could, therefore, remove resources from the



process of stopping speech and *predictively* allocate those resources to the process of retrieving the player's name, thus performing the latter task more efficiently (see Tydgat et al., 2011). Forward-model predictions could thus affect how quickly the commentator stops speaking, and could do so very rapidly.

In the same way, in our SELF condition, the cued participant sends a stop signal to the articulators. The participant also intends to name the target picture, and therefore sends a command to the production implementer. At the same time, a copy of the command is sent to a forward production model that computes a prediction that a word will be produced. At this point, the participant has not completed the process of stopping and is therefore still naming the initial picture. The prediction that the target word will be produced triggers the (predictive) reallocation of resources from the process of stopping to the process of retrieving the target word, thus delaying the stopping process. In addition, the cued participant retrieves the target picture's name using the production implementer. This process takes up resources and further delays the process of stopping speech, thus increasing the tendency to complete the initial word.

When a speaker is planning an utterance, the predictions generated by the forward model are always accompanied by the activation of representations within the production implementer, which in turn normally leads to articulation. Crucially, according to Pickering and Garrod's (2013) account, forward-model predictions can be computed for another speaker's upcoming utterance as well (see Section 1). Note that, while forward production models are production mechanisms, and can affect ongoing language production, they can do so without activating the production implementer. In fact, according to Pickering and Garrod the production implementer is not required for prediction of other people's utterances and the activation in D'Ausilio et al. (2011) could potentially be incidental, rather than causally involved in prediction.

Now imagine a situation in which our sports commentator is assisted by a (male) partner. While describing some background details, she realizes that her partner has noticed the action. She predicts that he is about to speak. However, she does not retrieve the semantics, syntax, or phonology of her partner's utterance and does not therefore have to take resources away from her own production. But because she predicts that her partner is about to speak using the same mechanism that she would use to predict that she is about to speak, she might nevertheless take some resources away from stopping (because this has proved effective in the past). Note that this might not be an effective strategy, as it might be argued that the commentator should try and stop as quickly as possible. Nevertheless, delaying stopping could be beneficial in other ways; for example, it might give her partner more time to get ready whatever he is planning to say.

In a similar way, the cued participant in our OTHER condition knows that her partner intends to name the target picture, and therefore forms a representation of his production command. A copy of the command is sent to the forward production model and, as in the SELF condition, the prediction that a word will be produced triggers the reallocation of resources, away from the process of stopping. Unlike in the SELF condition, however, the participant does not use the representation of her partner's production command to drive retrieval of linguistic representations within the implementer. Therefore, the (partner's) naming of the target picture does not interfere with the process of stopping speech. This might explain why the tendency to complete the initial word was weaker in the OTHER than in the SELF condition.

In conclusion, our results indicate (i) that some mechanisms used during production are implicated in prediction; (ii) that production mechanisms are not as strongly activated when speakers predict others as when they prepare their own utterances. Our results are therefore compatible with both versions of the overlapping mechanisms account, that is with

a version in which the mechanisms overlapping between production and prediction are linguistic in nature (i.e., a subset of the mechanisms used during language production, whether forward model computations or the retrieval of linguistic representations), and with a version in which the mechanisms overlapping between production and prediction are general purpose and common to the preparation of non-verbal actions.

#### **4.2 Prediction and Between-Speaker Coordination in Dialogue**

What is the relevance of these results for our understanding of the coordination that takes place between speakers in natural conversations? Clearly, our task is very different from natural conversation. First, speaker *A* and speaker *B* produce two completely unrelated utterances. Second, the moment at which the speaker-switch occurs, and the direction of the switch, are fixed and determined by the experimenter. Third, the experimental conditions were particularly favorable for prediction: *A* could see what picture *B* was going to name, and she knew, because of the instructions, that *B* was about to name it. (This contrasts with the situation faced by the sport commentator, who might be much more uncertain about her partners upcoming action).

In addition, it is possible that participants in the OTHER condition developed a tendency to attend to the target picture even on trials on which they did not have to name it. The cued participant does the same thing in the NO and OTHER conditions, except that the cued participant names some target pictures in the OTHER condition, and these pictures of course have the same characteristics as the pictures that their partner names on the other trials. This aspect of the task might enhance the activation of production processes related to the target pictures, and in turn cause the cued participant to complete the names of the initial pictures more often in the OTHER condition than the NO condition. If so, it may be the case that people are affected by the fact that their partner is about to speak more under conditions in which they sometimes have to speak themselves.

Note that, because the proportion of change trials was very low in our experiment, naming the target picture on half of those trials is unlikely, on its own, to account for the increased tendency to complete the initial word in the OTHER compared to the NO condition. Moreover, participants in the OTHER condition never had to name an initial and a target picture on the same trial. Nevertheless, a future study could test whether participants who always name either the initial or the target pictures (i.e., they never switch roles) would tend to complete the initial word to the same or to a smaller extent as they did in the OTHER condition in this study.

Appendix B reports that the interval between the first and second picture names was longer in the OTHER compared to the SELF condition. This finding suggests that within-speaker coordination was not as successful as between-speaker coordination. It implies that between-speaker coordination is not identical to within-speaker coordination, and that predicting that another is about to speak is only one of the processes that support the coordination of utterances between speakers in conversation (Vesper, Butterfill, Knoblich, & Sebanz, 2010).

Despite such limitations, we suggest that our experiment provides evidence about some mechanisms that can be used in conversation (at least when conditions are favorable, and in conjunction with other mechanisms). Note that many natural conversations could support accurate predictions, because interlocutors can capitalize either on a long interactional history that leads to alignment (Pickering & Garrod, 2004) or on the highly formulaic nature of language in many activity types (e.g., purchasing an item in a shop, fixing an appointment at the doctor; Clark, 1996; Levinson, 1992). Moreover, production processes are particularly likely to be activated in natural conversations where interlocutors switch between the roles of speaker and listener all the time.

In addition, while we have only provided evidence that speakers can predict that their partner is going to speak, it is possible that similar mechanisms underlie the ability to predict what one's partner is going to say and when. Clearly, interlocutors would greatly benefit from the ability to predict (i) *that* their partner is going to speak; (ii) *what* their partner is going to say; (iii) *when* their partner is going to speak. Prediction (i) would allow them to decide whether to continue or stop speaking themselves. Prediction (ii) would help them prepare an adequate response to the current speaker's contribution, or to complete the speaker's utterance. Prediction (iii) would be useful for smooth turn-taking.

The idea that the computation of predictions underlies interlocutors' ability to coordinate their utterances is consistent with the *projection theory* of turn taking (Sacks, Schegloff, & Jefferson, 1974), whose central claim is that listeners predict possible completion points of a speaker's utterance as it unfolds, and that such completion points are where a change of speaker can potentially occur. Corpus studies have shown that inter-turn gaps tend to be very short, with mode offset of the distribution varying between 0 and 200 ms for a range of languages (Stivers et al., 2009). People take 500-700ms to start speaking in reaction to a cue, even if the material is pre-planned and only needs to be retrieved from memory (e.g., Ferreira, 1991). Thus, interlocutors must regularly predict turn endings, and plan their contribution in advance. De Ruiter and colleagues have indeed shown that listeners are able to reliably predict when a turn is going to end and that this ability is based on a prediction of what the speaker is going to say (De Ruiter, Mitterer, & Enfield, 2006; Magyari & De Ruiter, 2012).

More generally, the idea that processes involved in production might be recruited to predict and comprehend the utterances produced by one's interlocutor provides an explanation of how people are able to perform the joint activity of dialogue, as discussed by Pickering and Garrod (2013). They proposed that interlocutors who mutually predict one

another using production mechanisms eventually become coordinated, and that this straightforwardly explains the smoothness of turn-taking and the occurrence of cross-person contributions. In addition, prediction might underlie alignment of representations (Pickering & Garrod, 2004); and, vice versa, prediction tends to be successful because interlocutors become sufficiently similar during the conversation (so that they can predict what the other is about to say by predicting what they themselves would say next). However, these theoretical considerations are yet to be supported by experimental findings. We suggest that joint language tasks like the one employed in this study could be used to investigate, for example, whether speakers whose representations are aligned (e.g., similar speech rate, lexical choices) can coordinate better with one another (Gambi & Pickering, 2013).

## **5. Conclusion**

We showed that speakers can predict that another person is going to speak and that such predictions can affect how quickly they stop their own utterance. This indicates that predicting that another is about to speak makes use of some of the processes involved in preparing to speak. Future research should investigate whether such processes are general purpose or domain specific and how they could be used to support coordination between speakers in dialogue.

## **6. Acknowledgements**

We thank Rob Hartsuiker for helpful discussions and for making the target pictures available to us, Eddie Dubourg (Edinburgh University) for invaluable technical assistance, and Michael Stevens (Ghent University) for patiently discussing various statistical issues with us. Chiara Gambi was supported by a University of Edinburgh scholarship. Uschi Cop is

supported by an FWO (Fonds Wettenschappelijk Onderzoek) scholarship. Neither funding source influenced the work reported in this paper.

### References

- Altmann, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Audition. (Version 1.05) [Computer Software]. San Jose, CA: Adobe Systems Incorporated.
- Retrieved from [www.adobe.com](http://www.adobe.com)
- Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457-474.
- Bates, D., Maechler, M., & Dai, B. (2008). Lme4: Linear mixed-effects models using S4 classes. (Version 0.999375-28 ).
- Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412.
- Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, 39(3), 173-194.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 4.6.22). <http://www.praat.org/>.
- Clark, H. H. (1996). *Using language*. Cambridge, U.K.: Cambridge University Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (Third ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

- Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 162-175.
- D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882-887.
- D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., & Craighero, L. (2011). Tongue corticospinal modulation during attended verbal stimuli: Priming and coarticulation effects. *Neuropsychologia*, 49(13), 3670-3676.
- De Bruijn, E. R., Miedl, S. F., & Bekkering, H. (2008). Fast responders have blinders on: ERP correlates of response inhibition in competition. *Cortex*, 44(5), 580-586.
- De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), 515-535.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, 121(3), 226-239.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121.
- E-Prime. (Version 2.0) [Computer Software]. Pittsburgh, PA: Psychology Software Tools. Retrieved from [www.pstnet.com](http://www.pstnet.com)
- Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, 15(2), 399-402.



- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161.
- Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75-84.
- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, 30(2), 210-233.
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans in action observation. *Nature*, 424(6950), 769-771.
- Gambi, C., & Pickering, M. J. (2013). Talking to each other and talking together: Joint language tasks and degrees of interactivity. *Behavioral and Brain Sciences*, 36(4), 423-424.
- Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage*, 36(Supplement 2), T22-T32.
- Griffin, Z. M. (2003). A reversed word length effect in coordinating the preparation and articulation of words in speaking. *Psychonomic Bulletin & Review*, 10(3), 603-609.
- Guediche, S., Holt, L. L., Laurent, P., Lim, S. J., & Fiez, J. A. (2014). Evidence for cerebellar contributions to adaptive plasticity in speech perception. *Cerebral Cortex*, doi: 10.1093/cercor/bht428.
- Hartsuiker, R. J., Catchpole, C. M., De Jong, N. H., & Pickering, M. J. (2008). Concurrent processing of words and their replacements during speech. *Cognition*, 108(3), 601-607.

- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the Perceptual Loop Theory. *Cognitive Psychology*, 42(2), 113-157.
- Hartsuiker, R. J., Pickering, M. J., & De Jong, N. H. (2005). Semantic and phonological context effects in speech error repair. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 921-932.
- Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, 22(7), 1114-1139.
- Hosemann, J., Herrmann, A., Steinbach, M., Bornkessel-Schlesewsky, I., & Schlewsky, M. (2013). Lexical prediction via forward models: N400 evidence from German sign language. *Neuropsychologia*, 51(11), 2224-2237.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151-171.
- Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, 9(4), 304-313.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.
- Kilner, J. M., Varga, C., Duval, S., Blakemore, S.-J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, 7(12), 1299-1301.

Klauer, K. C., Herfordt, J., & Voss, A. (2008). Social presence effects on the stroop task:

Boundary conditions and an alternative account. *Journal of Experimental Social Psychology*, 44(2), 469-476.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action:

Theory and Data. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 59-101). Burlington: Academic Press.

Knoblich, G., & Jordan, G. S. (2003). Action coordination in groups and individuals:

Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 1006-1016.

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Cerebellar contribution to the prediction of self-initiated sounds. *Cortex*, 49(9), 2449-2461.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead:

Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). Oxford: Oxford University Press.

Lesage, E., Morgan, B. E., Olson, A. C., Meyer, A. S., & Miall, R. C. (2012). Cerebellar

rTMS disrupts predictive language processing. *Current Biology*, 22(18), R794-R795.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1-75.

Levinson, S. C. (1992). Activity types and language. In P. Drew & J. Heritage (Eds.), *Talk at work: Interaction in institutional settings* (pp. 66-100). Cambridge: Cambridge

University Press.

- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126-1177.
- Magyari, L., & De Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology*, 3. doi: 10.3389/fpsyg.2012.00376
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake - but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), 843-847.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1), 82-91.
- Meyer, A. S., Belke, E., Häcker, C., & Mortensen, L. (2007). Use of word length information in utterance planning. *Journal of Memory and Language*, 57(2), 210-231.
- Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, 66(2), B25-B33.
- Mostofsky, S. H., & Simmonds, D. J. (2008). Response inhibition and response selection: Two sides of the same coin. *Journal of Cognitive Neuroscience*, 20(5), 751-761.
- Nooteboom, S. G. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in linguistic performance* (pp. 87-95). New York: Academic Press.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, 38(8), 1147-1160.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169-226.
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105-110.

- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-392. doi: <http://dx.doi.org/10.1017/S0140525X12001495>
- Pickering, M.J. & Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Frontiers in Human Neuroscience*, 8:132. doi: 10.3389/fnhum.2014.00132
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413-425.
- R. (Version 2.7.2) [Computer program]. Vienna, Austria: R Development Core Team, <http://www.R-project.org>.
- Ramnani, R., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience*, 7(1), 85-90.
- Riès, S., Janssen, N., Dufau, S., Alario, F. X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, 23(6), 1419-1436.
- Roy, A. C., Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology - Paris*, 102, 101-105.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696-735.
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10(2), 70-76.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353-367.

- Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in coacting individuals. *Journal of Cognitive Neuroscience, 18*(5), 859-870.
- Sebanz, N., Rebbechi, D., Knoblich, G., Prinz, W., & Frith, C. D. (2007). Is it really my turn? An event-related fMRI study of task sharing. *Social Neuroscience, 2*(2), 81-95.
- Serlin, R. C., & Levin, J. R. (1985). Teaching how to derive directly interpretable coding schemes for multiple regression analysis. *Journal of Educational Statistics, 10*(3), 223-238.
- Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition, 108*(3), 837-842.
- Snodgrass, J. G., & VanderWart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174-215.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., . . . Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*(26), 10587–10592.
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience, 25*(7), 1020-1036.
- Tsai, C.-C., Kuo, W.-J., Jing, J.-T., Hung, D. L., & Tzeng, O. J. L. (2006). A common coding framework in self-other interaction: Evidence from joint action task. *Experimental Brain Research, 175*(2), 353-362.

- Tsai, C.-C., Kuo, W.-J., Hung, D. L., & Tzeng, O. J. L. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, 20(11), 2015-2024.
- Tydgat, I., Stevens, M., Hartsuiker, R. J., & Pickering, M. J. (2011). Deciding where to stop speaking. *Journal of Memory and Language*, 64(4), 359-380.
- Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176-190.
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8), 998-1003.
- Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, 19(4), 403-440.
- Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, 41(8), 989-994.
- Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology*, 2(2), 147-172.
- West, S. G., Aiken, L. S., & Krull, J. L. (1996). Experimental personality designs: Analyzing categorical by continuous variable interactions. *Journal of Personality*, 64(1), 1-48.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460-473.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209-216.

Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M., Shtyrov, Y., & Mäkelä, J. P. (2014). Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. *Cerebral Cortex*, doi: 10.1093/cercor/bht351.

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, 107(2), 592-597.



*Appendix A*

Items Used on Change Trials

INSERT TABLE A1 HERE

*Appendix B*

## Supplementary Analyses

For the analyses of Target Onset and Interval (in the SELF and OTHER conditions), we excluded all trials in which the initial or the target word was named incorrectly or produced disfluently (14.7% in SELF, 15.0% in OTHER). For Initial Onset and Interval, we also excluded all trials with skipped initial responses (but we included interrupted responses, as we were interested in the effect of Response Type; see Table 2). In the OTHER condition, it was possible to have negative values for Interval (overlap between the two participants' responses). Such cases (6.5%) were also excluded from the analyses of Target Onset and Interval in the OTHER condition for the sake of comparability between conditions. For each condition separately, we then removed all trials that were more than 2.5 SD from the grand mean or more than 3 SD from the by-subject mean from the analyses (initial onset: 3.0% in SELF, 2.7% in OTHER, 3.0% in NO; target onset: 3.4% in SELF, 2.9% in OTHER; initial-target interval: 4.9% in SELF, 2.7% in OTHER).

The starting point was the model including the factors Condition, Degradation, Response Type, all the possible two-way interactions, and the three-way interaction. We set the SELF condition as the reference level, and we defined two contrasts, one comparing the weighted average of OTHER and NO to SELF (Condition1), and the other comparing OTHER against NO (Condition2). We used weighted contrast coding for all predictors. So, for example, the contrast for the two-level factor Response Type was not (-.5, .5), but rather  $(-.5 \cdot n_i / N; .5 \cdot n_c / N)$ , where  $n_i$  is the count of interrupted responses,  $n_c$  is the count of completed responses and  $N = n_i + n_c$  (completed responses were taken as the reference level). In this way, the intercept corresponds to the weighted grand mean and the estimates

for the main effects are equal to twice the difference between levels of the corresponding factor.

INSERT TABLE B1 ABOUT HERE

INSERT TABLE B2 ABOUT HERE

The onset of the initial picture name (see Tables B1 and B2) was delayed for interrupted initials compared to completed initials. There was also some indication of the effect being larger in the SELF (756 vs. 700 ms) and OTHER (755 vs. 716 ms) conditions than in the NO condition (703 vs. 705 ms), as indicated by a Response Type by Condition interaction. Finally, the effect of Response Type was larger before intact than before degraded targets in the SELF (85 vs. 23 ms) and the OTHER (60 vs. 15 ms) conditions (see significant Response Type by Degradation interaction and the marginal three-way interaction of Response Type, Degradation, and Condition). Hartsuiker et al. (2008) also reported longer onset times for interrupted than for completed initials (though only in their Experiment 2), and suggested that it was due to the stopping process being more likely to stop word-internally when the initial word is initiated later.

INSERT TABLE B3 ABOUT HERE

INSERT TABLE B4 ABOUT HERE

The Interval between the offset of the initial name and the onset of the target name (in the SELF and OTHER conditions) was longer before degraded than before intact targets and longer in the OTHER than in the SELF condition (see Tables B3 and B4). The three-way interaction of Response Type, Degradation, and Condition was marginal, suggesting that the effect of Degradation was larger before interrupted (79 ms) than before completed (11 ms) responses in the SELF condition but not in the OTHER condition (14 vs. 69 ms). Hartsuiker et al. (2008) reported a non-significant trend in the same direction in their Experiment 1.

In addition, in the SELF condition we observed a numerical trend for longer intervals after interrupted (146 ms) than after completed (91 ms) initial names. Hartsuiker et al. (2008) reported a significant difference in the same direction. According to them, this is because participants have more time to plan the target name while still articulating the initial name when they complete than when they do not. The difference was smaller in our experiment (55 ms) than in Hartsuiker et al. (150 ms), perhaps because intervals after interruptions were shorter in our experiment than theirs (146 vs. 216 ms). This might depend, in turn, on our participants' reduced propensity to interrupt (see Discussion).

INSERT TABLE B5 ABOUT HERE

INSERT TABLE B6 ABOUT HERE

Finally, the onset of the target picture name was delayed for degraded with respect to intact targets in both the SELF (809 vs. 777ms) and the OTHER condition (921 vs. 868ms), indicating that the manipulation was effective in both conditions (see Tables B5 and B6). Target onset latencies were also significantly longer in the OTHER (894 ms) than in the

SELF condition (792 ms). Target onsets varied as a function of initial Response Type, but differently in the two conditions. In the SELF conditions, latencies were much longer after skipped initials (878 ms) than after completed (785 ms) or interrupted (760 ms) initials. In the OTHER condition, instead, the target was named faster when the initial was interrupted (852 ms) or skipped (844 ms) than when it was completed (924 ms).